Squaring the circle of high-throughput chemistry

Andrew Lemon, IDBS, 2 Occam Court, Surrey Research Park, Guildford, UK GU2 7QB; e-mail: alemon@id-bs.com

The American Chemical Society recently held its 225th national meeting in New Orleans, USA (23–27 March 2003). The conference attracted ~15,000 delegates, ranging from students and academic experts to industry professionals from all kinds of chemically related disciplines, including discovery. The division of chemical information provided a key indicator of the direction that information professionals in discovery are following and how they are attempting to solve the problems surrounding new drug discovery.

Predicting toxicology

Scott Lee of MDL Information Systems (http://www.mdl.com) chaired an interesting session addressing the problem of toxicology prediction, to enable directed synthesis and *in silico* screening of potential hits earlier in the discovery process. The computational chemistry community has shifted its emphasis from the prediction and identification of drug molecules from large sets of candidates to the more tractable problem of predicting ADME/Tox properties of potential drugs.

Julie Penzotti described how Rational Discovery (http://www.rationaldiscovery.com) has created an integrated framework that connects screening and chemical data in computational models to support the prediction of physiochemical properties. The work of Rational Discovery is based on a composite model combining multiple single models of various types (including neural networks, regressions and decision trees) and a heuristic algorithm for picking the most appropriate combination of models to make a prediction.

Penzotti then described the application of these techniques to the prediction of oral toxicity in mice and blood–brain penetration, both of which involved training sets of <250 compounds. Predictions based on the composite models proved to be 10% more accurate than those based on individual models. Emphasis was put on the use of these techniques to direct synthesis and screening by homing in on key molecules. However, some of the audience was not convinced that the statistics demonstrated the benefit of using composite models.

Exchanging toxicological data

Chihae Yang of Leadscope (http://www.leadscope.com) described some of the problems of studying toxicity, especially ecotoxicity, with regard to the accessibility to data, much of which remains proprietary. Furthermore, the value of data in the public domain is reduced by the diversity of protocols used in its collection. A combination of mechanisms and factors work together to generate toxicity, and this complexity is further compounded when applied to whole ecosystems. Yang described some of the problems faced in the normalization of data collected from many publicly available sources and based on different methodologies.

As a result of these difficulties, Leadscope has initiated an effort to define an XML format called ToxML for the standardization of information exchange, but mainstream acceptance for this type of effort can take many years. In some respects, this description of the problems associated with data normalization and of the subjectivity of the process suggests that the extensive interpretations made on the data sets used might be unsafe.

Michael Berthold of the Data Analysis Research Lab at Tripos (http://www.tripos.com) presented their study on fragment mining in large datasets. Using market–basket analysis techniques, they have identified pharmacophores that are responsible for activity. It was refreshing to see a group looking outside of the norm to apply new techniques to the same old problems.

Increasingly, computational techniques are providing real benefit to ADME and toxicological assessment when designing experiments and compiling libraries of potential candidates. As these techniques are refined and assessed, their predictive capabilities will become a useful tool for the medicinal chemist, contributing to the overall discovery process.

High-throughput chemistry

David Evans of MDL Information Systems eloquently introduced this session, which combined several aspects of high-throughput chemistry, logistics, synthesis and analytics. Operating in a high-throughput *modus operandi* clearly generates several problems, which were highlighted during the session.

Greg Landrum of Rational Discovery described the use of their predictive framework in addressing the problem of directing combinatorial synthesis using predictive models. Landrum asserted that rapid 2D models provide equivalent predictive and screening power to the more intensive and involved 3D techniques, making them

1359-6446/03/$ – see front matter PII: S1359-6446(03)02748-X

more applicable in a high-throughput environment. This was clearly controversial to the proponents of 3D methodologies.

Several talks described the overall process from design through virtual screening to reagent acquisition, synthesis and analytics. John Brohan of Trader's Micro (http://www.tradersmicro.com) and Rejean Fortin of Merck Frosst Canada (http://www.merckfrosst.ca) presented an application of high-throughput analytics and synthesis using the GeminiChemistry system, another important issue in high-throughput chromatography (HTC). This was further emphasized by Mike McBrien of Advanced Chemistry Development Laboratories (http://www.acdlabs.com), who described their work on the prediction of retention times in HTC.

An Integrated discovery platform

There were three talks from ArQule (http://www.arqule.com), which together highlighted some important principles in managing and optimizing library size, number of leads and drug-likeness, and in the minimization of ADME and toxicity attrition, while also improving the overall efficiency of much of the discovery process.

David Hartsough presented a vision of an integrated discovery platform that spans design, synthesis, logistics, inventory and analytics. The process is tracked and audited, providing more visibility and accountability to comprising elements. This is essential for managing the process and ensuring that bottlenecks are identified by actual cause and effect, not hunch and accusation, leading to a more open culture of efficiency. Although this is an adjustment for many scientists, it is hard to resist when backed by hard facts and reports.

Hartsough described how ArQule has created an integrated IT system that connects each part of the business. He explained the benefits of data accessibility through effective reporting and auditing, answering questions such as:

- How long did this array of compounds take from design to finished plate?
- What stage is this array at?
- What have we achieved this month, quarter, year?

Another business benefit is a much greater understanding of the precise costs of libraries than was previously possible, enabling organizations to tune resources allocated to elements of the process, thereby improving the overall efficiency.

Brian Deneau presented the ArQule integrated analytics process that enables the efficient quality control of libraries in a factory-orientated procedure, and has moved ArQule from a 10,000-samples-per-year, four-person operation to an impressive 500,000-samples-per-year, eight-person operation. Integration enables ArQule to produce metrics for the number of plates and samples processed by time slot and project. Through analysis of the components of the process, the system detects how much time is spent on data submission, analysis and review.

Daniel Gschwend presented the third element of the ArQule process: a reagent acquisition solution that supports efficient ordering and inventory. Access to important metrics for picking reagent suppliers was essential, and included factors such as price, purity and availability. The turnaround time of suppliers, the quality of received goods and the success of ordered libraries were measured to assess the ordering procedure. As a result, ArQule has improved contract negotiation with suppliers, proving that the benefits of an effective information management system can be far reaching.

Business analytics

The ArQule vision indicates a new professionalism in management of the discovery process that streamlines and directs information from target to final drug. This approach places increasing importance on collecting and learning from each piece of data and reinvesting this knowledge into making improved candidates more efficiently and at a lower cost. The demand from the industry for this type of business analytics will rise to control the ever-increasing cost of the discovery process. Improvements in many parts of the process, from toxicology and ADME property prediction to high-throughput analytics, integrated through a modern informatics infrastructure, are shaping this vision. However, high quality and consistent data are crucial to all of this, and prediction models that are built on a solid foundation of good data.

Bright light of the future

Like many new techniques, combinatorial chemistry burned brightly with promises of a factory for new drugs. Now, in the glowing embers, we see its true value and contribution to discovery. Perhaps the increased intensity of the high throughput operation in screening and combinatorial chemistry will generate benefits for the overall process.